

THE MICROSCOPIC APPROACH TO $\mathcal{N} = 1$ SUPER YANG-MILLS THEORIES

FRANK FERRARI

*Service de Physique Théorique et Mathématique
Université Libre de Bruxelles and International Solvay Institutes
Campus de la Plaine, CP 231, B-1050 Bruxelles, Belgique
frank.ferrari@ulb.ac.be*

We give a brief account of the recent progresses in super Yang-Mills theories based in particular on the application of Nekrasov's instanton technology to the case of $\mathcal{N} = 1$ supersymmetry. We have developed a first-principle formalism from which any chiral observable in the theory can be computed, including in strongly coupled confining vacua. The correlators are first expressed in terms of some external variables as sums over colored partitions. The external variables are then fixed to their physical values by extremizing the microscopic quantum superpotential. Remarkably, the results can be shown to coincide with the Dijkgraaf-Vafa matrix model approach, which uses a totally different mathematical framework. These results clarify many important properties of $\mathcal{N} = 1$ theories, related in particular to generalized Konishi anomaly equations and to Veneziano-Yankielowicz terms in the glueball superpotentials. The proof of the equivalence between the formalisms based on colored partitions and on matrices is also a proof of the open/closed string duality in the chiral sector of the theories.

Keywords: Supersymmetric gauge theories; Gauge-gravity correspondence; Non-perturbative effects.

PACS numbers: 11.30.Pb, 11.15.Tk, 11.25.Tq

1. Introduction

The study of non-perturbative properties of supersymmetric gauge theories, by making extensive use of their special properties like holomorphy, is by now a rather mature field of research, and a great wealth of results have been obtained over the last 15 years. Many techniques and ideas have been used (electric-magnetic duality, integrable systems, mirror symmetry, brane engineering...), producing an extremely rich and consistent picture of the strongly coupled regime of super Yang-Mills models. Until recently, the most general approach was based on an application of the open/closed string duality to the holomorphic, or chiral, sector of the theory.¹ The result of this approach is most elegantly encoded in a planar matrix model and a suitable glueball superpotential.²

Many attempts have been made to provide gauge theoretic justifications of the matrix model results. For example, in Ref. 3, the present author provided a proof of the matrix model conjecture (in the so-called “one-cut” case) based on the

assumptions of confinement and the Intriligator-Leigh-Seiberg linearity principle. These assumptions are of course very natural, but extremely difficult to justify from first principles. Other works focused on the direct calculation of the glueball superpotential⁴ or the use of generalized Konishi anomaly equations.⁵ However, these studies were limited to a perturbative analysis, whereas the main interest in the matrix model is its ability to provide exact *non*-perturbative results. The conclusion is that in spite of their great interest in improving our intuitive understanding, all the above-mentioned approaches fall short in providing valid proofs of the results.

Our aim in this talk is to present a *first principle, microscopic* approach to the non-perturbative dynamics of $\mathcal{N} = 1$ gauge theories in the chiral sector, which amounts to computing directly the relevant path integrals without making any approximation or assumption.^{6–8} The general observable that we compute is the expectation value of an arbitrary chiral operator,

$$\langle \mathcal{O} \rangle(\mathbf{g}, q). \quad (1)$$

Such expectation values cannot be corrected in perturbation theory (which also explains why any perturbative argument that aims at computing them is doomed to fail), but they do get very non-trivial and interesting non-perturbative corrections. They depend in general on the various parameters in the gauge theory action. The gauge coupling constant enters in the quantum theory through the instanton factor q . There are also various couplings in the tree-level superpotential and prepotential (see below), that we have denoted collectively by \mathbf{g} in Eq. (1). A typical example of a chiral correlator is the gluino condensate in the pure $\mathcal{N} = 1$ gauge theory, which is proportional to a fractional power of q , $\langle \text{tr } \lambda \lambda \rangle = q^{1/N}$. From the knowledge of the expectation values (1), one can derive the quantum vacuum structure and the phases of general $\mathcal{N} = 1$ theories,^{11–14} the Seiberg-Witten solution of the theories with extended supersymmetry,¹⁵ and actually all the known exact results in supersymmetric gauge theories.

Our microscopic approach is based on Nekrasov's instanton technology.^{16–20} This was originally developed for theories with $\mathcal{N} = 2$ supersymmetry, and our main contribution is to extend it to the $\mathcal{N} = 1$ case (an important early work in this direction was given in Ref. 21). Our results can be seen as the open string solution of the model, since we start from the gauge theory action. The solution is expressed in terms of averages over colored partitions that can be explicitly evaluated. This description is mathematically very different from the Dijkgraaf-Vafa recipe² that uses averages over hermitian matrices. Our proof in Ref. 8, reviewed below, that the two formalisms yield exactly the same results for the physical correlators (1) provides a full justification of the matrix model and equivalently a proof of the open/closed string duality for the chiral sector of the $\mathcal{N} = 1$ gauge theories.

An interesting application of our results is to revisit the perturbatively-derived anomaly equations of Ref. 5 and see if and in what sense they remain true at the non-perturbative level. As we shall explain, the anomaly equations are non-trivial

dynamical relations from the microscopic point of view (whereas they correspond to rather trivial identities, the loop equations, in the matrix model framework). An interesting conceptual result is that, once these equations have been understood at the non-perturbative level, the full solution of the model follows.^{9,10} The remaining ambiguities, that are related to the choice of particular Veneziano-Yankielowicz terms in the glueball superpotentials,⁵ turn out to be completely fixed by general consistency conditions.

The plan of the talk is as follows. In the next Section, we present the model on which we focus and explain the main results. In Section 3, we discuss the generalized anomaly equations, emphasizing the important gap between a perturbative study and the non-perturbative analysis required to compute the correlators (1). The power of the non-perturbative anomaly equations is fully revealed in the consistency theorem of Ref. 9 and 10 (Theorem 3.2), which we explain. In Section 4, we give more details on the microscopic formalism. Finally, we discuss possible generalizations and conclude in Section 5.

2. The Model and Sketch of the Main Result

We focus on the paradigmatic example of the $U(N)$ theory with one adjoint chiral superfield X . Let us note that there is no difficulty in considering a more general matter content. If needed, one can also integrate out the adjoint field by sending its mass to infinity at the end of the calculations. If we note W^α the super field strength that contains the gauge field and the gluino, the lagrangian takes the form

$$L = 2 \operatorname{Re} \int d^2\theta \mathcal{W} + \text{D-terms}, \quad (2)$$

with

$$\mathcal{W} = -\frac{1}{16\pi^2 N} \operatorname{Tr} t''(X) W^\alpha W_\alpha + N \operatorname{Tr} W(X). \quad (3)$$

The function $W(X)$ is an arbitrary polynomial tree-level superpotential, and $t''(X)$ is an arbitrary field-dependent polynomial gauge coupling. If $W = 0$, the theory has $\mathcal{N} = 2$ supersymmetry and $t(X)$ is an arbitrary tree-level prepotential. This generalized Seiberg-Witten model was studied in Ref. 20. When $W \neq 0$, the theory has only $\mathcal{N} = 1$ supersymmetry. Special cases are studied in Ref. 22. It turns out that the rôles of t and W are somehow interchanged in the microscopic and matrix model formalisms,⁸ which makes the consideration of the general theory (3) very natural. The case where $t'' = \ln q$ is a constant corresponds to the standard theory.

It is not difficult to show that any correlator (1) can be written as a sum of products of correlators of the basic variables

$$u_k = \operatorname{Tr} X^k, \quad v_k = -\frac{1}{16\pi^2} \operatorname{Tr} W^\alpha W_\alpha X^k \quad (4)$$

that generate the chiral ring. It is thus convenient to encode the solution of the

4 *Frank Ferrari*

theory into two generating functions

$$R(z) = \sum_{k \geq 0} \frac{\langle u_k \rangle}{z^{k+1}}, \quad S(z) = \sum_{k \geq 0} \frac{\langle v_k \rangle}{z^{k+1}}. \quad (5)$$

The solution derived in the microscopic formalism takes the following form. First one introduces averages over the ensemble of colored partitions (more details are given in Section 4), which physically label certain field configurations over which the path integrals localize. If we denote these averages with the symbol $[\]$, we compute

$$u_{k, \text{mic}}(\mathbf{a}) = \lim_{\epsilon \rightarrow 0} [\text{Tr } X^k]. \quad (6)$$

The parameter ϵ corresponds to a deformation of the gauge theory (the so-called Ω -background) that one must consider in order to define the measure on colored partitions. The result for the original gauge theory are obtained by taking the $\epsilon \rightarrow 0$ limit. The parameters $\mathbf{a} = (a_1, \dots, a_N)$ correspond to arbitrary boundary conditions at infinity for the adjoint field X ,

$$X_\infty = \text{diag } \mathbf{a}. \quad (7)$$

The importance of these boundary conditions will be explained in more details in Section 4. It is also possible to express $[v_k]$ in terms of the $[u_{k'}]$,^{21,7}

$$v_{k, \text{mic}}(\mathbf{a}) = \frac{N}{(k+1)(k+2)} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon^2} ([\text{Tr } W(X) \text{Tr } X^{k+2}] - [\text{Tr } W(X)][\text{Tr } X^{k+2}]). \quad (8)$$

Generating functions $R_{\text{mic}}(z; \mathbf{a})$ and $S_{\text{mic}}(z; \mathbf{a})$ can then be defined following Eq. (5). To obtain the physical correlators, one must fix \mathbf{a} to special values, that are obtained by extremizing the microscopic superpotential⁶

$$W_{\text{mic}}(\mathbf{a}) = [\text{Tr } W(X)]. \quad (9)$$

The set of solutions $\mathbf{a} = \mathbf{a}^*$ to the equations

$$dW_{\text{mic}}(\mathbf{a} = \mathbf{a}^*) = 0 \quad (10)$$

turns out to be in one-to-one correspondence with the full set of quantum vacua of the theory.⁶ The physical generating functions are then given by

$$R(z) = R_{\text{mic}}(z; \mathbf{a}^*), \quad S(z) = S_{\text{mic}}(z; \mathbf{a}^*). \quad (11)$$

It is interesting to compare the above formalism with the Dijkgraaf-Vafa matrix model formalism. There, one computes averages, that we shall denote by $\langle\langle \ \rangle\rangle$, over hermitian matrices. The basic identity relates the glueball operators v_k to matrix model averages,

$$v_{k, \text{MM}}(\mathbf{s}) = N \lim_{\epsilon \rightarrow 0} \epsilon \langle\langle \text{Tr } X^k \rangle\rangle. \quad (12)$$

The parameter ϵ (not to be confused with ϵ in the microscopic formalism) is related to the size of the matrices in the matrix model, $\epsilon \sim 1/n$, and thus we take the planar limit in Eq. (12) (although the number of colors N in the gauge theory is fixed and

finite). The arbitrary parameters $\mathbf{s} = (s_1, \dots, s_r)$ are called the filling fractions. Together with the integer r , $1 \leq r \leq \deg W'$, they label the general solution of the matrix model. The index “MM” in (12) emphasizes the fact that this is a matrix model average, and it should not be confused for example with $v_{k, \text{mic}}$ defined in Eq. (8). A formula for $u_{k, \text{MM}}(\mathbf{s})$ can also be given, which is formally similar to the right hand side of Eq. (8) but with matrix averages replacing averages over colored partitions and t'' replacing W (see Eq. (2.30) in Ref. 8). To $u_{k, \text{MM}}$ and $v_{k, \text{MM}}$ correspond generating functions $R_{\text{MM}}(z; \mathbf{s})$ and $S_{\text{MM}}(z; \mathbf{s})$ that have been studied extensively in the literature. Note that *they do not coincide* with the generating functions $R_{\text{mic}}(z; \mathbf{a})$ and $S_{\text{mic}}(z; \mathbf{a})$ of the microscopic formalism; they depend on different variables \mathbf{a} and \mathbf{s} , and they have in general different analytic structures (for example, it is well-known that R_{MM} and S_{MM} are two-valued, algebraic functions of z defined on a hyperelliptic curve; on the other hand, S_{mic} is infinitely multi-valued and thus not algebraic⁸). A basic conjecture of the matrix model formalism² is that the physical correlation functions are obtained for certain values of the filling fractions \mathbf{s} , that correspond, for each value of r , to the critical points $\mathbf{s} = \mathbf{s}^*$ of a suitable glueball superpotential $W_{\text{glue}}^{(r)}(\mathbf{s})$,

$$dW_{\text{glue}}^{(r)}(\mathbf{s} = \mathbf{s}^*) = 0, \quad 1 \leq r \leq \deg W'. \quad (13)$$

We can now state our main result. First, the set of solutions of the $\deg W'$ equations (13) are in one-to-one correspondence with the set of solutions of the *single* equation (10).⁶ Taking into account this correspondence, one can then show that⁸

$$\boxed{R_{\text{mic}}(z; \mathbf{a}^*) = R_{\text{MM}}(z; \mathbf{s}^*), \quad S_{\text{mic}}(z; \mathbf{a}^*) = S_{\text{MM}}(z; \mathbf{s}^*)}. \quad (14)$$

These fundamental identities are equivalent to the open/closed string duality in our case. They imply that when both formalisms are taken on-shell (i.e. when $\mathbf{a} = \mathbf{a}^*$ and $\mathbf{s} = \mathbf{s}^*$), then the generating functions computed using colored partitions and matrices coincide. They also provide the full justification of the matrix model recipe, since the microscopic formalism is a first-principle approach and the identification in Eq. (11) with the physical gauge theory correlators follows from the basic rules of QFT.

3. Non-Perturbative Anomalies and Consistency Conditions

In this Section, we are going to revisit the approach advocated in Ref. 5, which is based on the study of the gauge theory equations of motion, but now from a non-perturbative point of view. For simplicity, we limit the discussion to the standard case for which t'' is a constant.

3.1. The classical picture

To understand the nature of the reasoning and of the consistency conditions we want to use, it is useful to start by analysing the “trivial” case of the classical theory. If

6 *Frank Ferrari*

we note $d = \deg W'$, the classical equations of motion take the simple form

$$W'(X) = 0 \propto \prod_{i=1}^d (X - x_i), \quad (15)$$

which yields

$$R(z) = \sum_{i=1}^d \frac{N_i}{z - x_i}. \quad (16)$$

The positive integers N_i label the classical vacua $|N_1, \dots, N_d\rangle$ of the theory and correspond to the number of eigenvalues of the matrix X that are equal to x_i . The gauge symmetry breaking pattern is $U(N_1) \times \dots \times U(N_d)$. An interesting question to ask is the following: can we write the equations of motion (15) in terms of gauge invariant operators only? This is essential in view of possible generalizations to the quantum theory. A complete set of relations on gauge invariant operators derived from (15) is given by

$$\text{Tr}(X^{n+1}W'(X)) = 0, \quad n \geq -1. \quad (17)$$

These equations are obtained by considering variations of the form $\delta X \sim X^{n+1}$, which are generated by the Virasoro-like operators

$$L_n = -X^{n+1} \frac{\delta}{\delta X}, \quad [L_n, L_m] = (n - m)L_{n+m}. \quad (18)$$

In terms of the generating function $R(z)$, it is straightforward to check by expanding at large z that the equations (17) are equivalent to the requirement that the product

$$W'(z)R(z) = N\Delta(z) \quad (19)$$

must be a polynomial. In other words, the most general solution to Eq. (17) is of the form

$$R(z) = \frac{N\Delta(z)}{W'(z)} = \sum_{i=1}^d \frac{c_i}{z - x_i}. \quad (20)$$

This is *not* the expected solution, because the c_i s can be arbitrary complex numbers, whereas to match the correct solution (16) they must be positive integers! So it would seem that the description in terms of gauge invariant operators (which is also what one gets in a closed string formalism, where only gauge invariant objects can be introduced) is missing something. Intuitively, the open strings (matrix) can be built from the closed strings (gauge invariant operators) *only when some quantization conditions are satisfied*.

The fundamental idea to implement these quantization conditions in a gauge invariant language^{9,10} is that since the number of colors N in the gauge theory is *finite*, then the $u_k = \text{Tr } X^k$ are not all independent, but there must exist polynomial relations of the form

$$u_{N+p} = P_p(u_1, \dots, u_N) \quad (21)$$

for all $p \geq 1$. The explicit form of the polynomials P_p can be easily found. It turns out that there is a simple but nice algebraic lemma¹⁰ that states that the relations (21) are consistent with Eq. (20) *if and only if the c_i are positive integers*.

At the quantum *perturbative* level,⁵ one can still study the consequence of the variations generated by the operator L_n in Eq. (18) in the path integral. The result reads⁵

$$-N \sum_{k \geq 0} g_k u_{n+k+1} + 2 \sum_{k_1+k_2=n} u_{k_1} v_{k_2} = 0, \quad (22)$$

where we are using the expansion $W'(z) = \sum_{k \geq 0} g_k z^k$. The first term in the left hand side of Eq. (22) corresponds to the classical contribution that we already had in Eq. (17) and the second term is a one-loop anomaly called a generalized Konishi anomaly. In the quantum theory non-trivial information also comes from considering the variations generated by the operators

$$J_n = \frac{W^2}{16\pi^2} \frac{\delta}{\delta X}, \quad [L_n, J_n] = (n-m)J_{n+m}, \quad [J_n, J_m] = 0, \quad (23)$$

which yield

$$-N \sum_{k \geq 0} g_k v_{n+k+1} + \sum_{k_1+k_2=n} v_{k_1} v_{k_2} = 0. \quad (24)$$

Now comes an important point that was apparently completely overlooked in the early literature on this subject. The equations (22) and (24) must be supplemented with the constraints (21). These constraints are automatically valid to all orders of perturbation theory. It is then not difficult to show that the only solutions to Eq. (22), (24) and (21) are purely classical, $S(z) = 0$ and $R(z)$ given by Eq. (16). In some sense, we have just rederived, in a very roundabout way using anomaly equations, the standard perturbative non-renormalization theorem for chiral operators.

Note that the anomaly equations are very similar to the loop equations of the matrix model. Actually, Eq. (24) precisely coincides with the loop equations for the $\text{Tr } X^k$ in the planar limit, and this is how the relation (12) was explained in Ref. 5. However, we now see that there is a fundamental difference between the gauge theory and the planar matrix model. *In the planar matrix model, since the size of the matrix is infinite, all the variables that enter the loop equations are independent, and the most general solution is labeled by filling fractions. In the gauge theory, N is finite and there are constraints (21).*

3.2. The non-perturbative anomaly theorem

A direct consequence of the above analysis is that the anomaly equations must be quantum corrected. Otherwise the correlators would be purely classical! Quantum corrections can be a priori fairly general, the only obvious constraints coming from global symmetries. For example, Eq. (22) is replaced in the quantum theory by an

8 *Frank Ferrari*

equation of the general form

$$-N \sum_{k \geq 0} g_k u_{n+k+1} + 2 \sum_{k_1+k_2=n} u_{k_1} v_{k_2} + \sum_{t \geq 1} q^t \left(\sum_{k \geq 0} g_k A_{n,k}^{(t)}(u_p) + \sum_{k' \geq 0} C_{n,k'}^{(t)}(u_p) v_{k'} \right) = 0. \quad (25)$$

Similarly, the generators L_n and J_n , Eq. (18) and (23), and the algebra they generate do get strong quantum corrections that are computed explicitly in Ref. 7 and 8 and that we describe briefly in Section 4.5.

What is the strongest “anomaly theorem” that can be expected to be valid at the non-perturbative level? The precise statement is as follows.^{9,8}

Theorem 3.1. (*Non-perturbative anomaly theorem*) *It is possible to absorb the non-perturbative quantum corrections in the anomaly equations by a suitable redefinition of the variables that enter the equations.*

For example, redefinitions of the variables u_k of the form

$$u_k \rightarrow u_k + \sum_{t \geq 1} q^t c_k^{(t)}(u_p) \quad (26)$$

are allowed by the global symmetries if $k \geq 2N$. This simply means that the operators u_k for $k \geq 2N$ are ambiguous in the quantum theory and a definition in terms of the basic variables must be given. This is very similar to the ambiguities in defining composite operators in ordinary QFTs. A precise definition requires the choice of a particular scheme, and we explain in Section 4.3 that this is exactly what happens here at the non-perturbative level (perturbatively, composite chiral operators are unambiguous). One possible and *perfectly consistent* definition is actually given by Eq. (21), but other choices, corresponding to quantum corrected relations of the form

$$u_{N+p} = \mathcal{P}_p(u_1, \dots, u_N, q), \quad (27)$$

are also possible. It is crucial to understand that the choice of relations (27) is *totally arbitrary* and *do not* correspond to quantum corrections to the chiral ring. Actually, the ring generated by the u_k and q (which is the sector of the chiral ring of zero R-charge) is the polynomial ring $\mathbb{C}[u_1, \dots, u_N, q]$. It is well-known that this ring do not admit non-trivial deformations that preserve commutativity.

The content of Theorem 3.1 is now clear. For generic definitions of the variables u_k , for example with the choice (21), the anomaly equations have complicated explicit quantum corrections as in Eq. (25). However, *there exists canonical definitions of the variables, of the form (27), that make the quantum corrections to the anomaly equations implicit*. It is explained in Section 4.3 that this canonical choice is related to a particular regularization of the instanton moduli space.

The full proof of Theorem 3.1 is given in Ref. 8 using the solution from the microscopic formalism. It is however interesting to make a brief comment on another possible route, that was suggested in a footnote in Ref. 5 and that was tried in the

literature. The idea is to make an ansatz for the possible quantum corrections to the operators L_n and J_n , for example

$$L_n \cdot u_m = -mu_{n+m} + \sum_{t \geq 1} q^t r_{n,m}^{(t)}(u_p), \quad (28)$$

and then try to use Wess-Zumino consistency conditions to constrain the form of the anomaly equations. However, this approach fails, because the ansatz (28) turns out to be wrong. This can be checked on the exact expressions for the quantum operators L_n in Ref. 8 (see Section 4.5). The problem is that Eq. (28) assumes that the action of the quantum L_n on a chiral operator is given by a polynomial expression. Perhaps surprisingly, the quantum corrections turn out to be much stronger.

3.3. The chiral ring consistency theorem

Let us now explain the main result of Ref. 9 and 10. As is well-known, the most general solution to the anomaly equations depends on a finite number of arbitrary parameters: the quantum versions of the constants c_i that we had at the classical level in Section 3.1 and also the matrix model filling fractions. The usual recipe, in the matrix model formalism, is to postulate that the c_i must be positive integers and that the filling fractions are fixed by extremizing the glueball superpotentials, Eq. (13). In the microscopic approach, this is justified by the identities (14), as we develop in Section 4. However, the philosophy of Ref. 9 and 10 is to show that the same result can be obtained from Theorem 3.1 by using algebraic consistency conditions only.

This is an interesting conceptual result. It shows in particular that the early point of view on this problem was erroneous. Originally, it was thought that the anomaly equation part was “easy” and that a full non-perturbative justification, based for example on equations like (28), could be worked out rather straightforwardly. On the other hand, it was thought that the fixing of the filling fractions (that correspond to the gluino condensates in the gauge theory) to their physical values would be extremely difficult to justify rigorously (the fact that the quantization of the c_i was also highly non-trivial was completely missed in early works). Now the chiral ring consistency theorem means that the situation is quite the opposite: once the non-perturbative anomaly theorem is proven (which turns out to require the full power of the microscopic approach, invalidating in particular (28)), the fixing of the filling fractions (and of the c_i) follow simply from consistency (and not from dynamics)!

The proof of the theorem¹⁰ is a generalized version of the classical analysis given in Section 3.1. The main point is that the anomaly equations yield an *infinite* set of constraints on a *finite* set of variables in the gauge theory, because N is finite (this is very unlike the planar matrix model). Clearly, the consistency of an infinite number of equations for a finite number of variables is not obvious at all, and indeed most solutions to the anomaly equations do not satisfy the requirements. The precise theorem (derived in Ref. 10 in the case of the theory with an arbitrary number of

quark flavors, from which the theory with only the adjoint field X can be obtained by integrating out the quarks) is as follows.

Theorem 3.2. (*Chiral ring consistency theorem*) *A general solution to the anomaly equations is not consistent with the existence of relations of the form (27) between the variables. Consistency can be achieved only for one particular choice of the polynomials \mathcal{P}_p in (27) and only when the c_i s are positive integers and for special values of the filling fractions. These special values correspond precisely to the critical points of the glueball superpotentials, as in (13), defined to include uniquely specified Veneziano-Yankielowicz terms.*

Note that the conditions on c_i and the filling fractions are known to be equivalent to the quantization conditions of the compact periods of the one-form $R_{\text{MM}}dz$, which is an elegant and powerful way to implement the constraints.

Let us finish this Section with a few additional qualitative comments. The “philosophy” of the theorem is that the open strings can be built from the closed strings only when some quantization conditions are satisfied in the closed string theory. This is similar to the quantization of the RR flux in $\text{AdS}_5 \times S^5$ for example. In our case, the conditions follow from the extremization of the flux superpotential (which can be identified with the glueball superpotential). Another comment is that in the microscopic formalism, the existence of relations of the form (27) is trivially implemented, since the operators are constructed from a $N \times N$ matrix X . This implies that non-trivial dynamical equations in the closed string/matrix model framework (which ensures the consistency of (27) as shown by Theorem 3.2) must be exchanged with trivial identities in the open string/microscopic formulation. This is exactly what is found.⁸ On the other hand, off-shell (i.e. valid for any \mathfrak{s}) identities in the closed string/matrix model formulation, like the generalized anomaly equations, correspond to highly non-trivial dynamical identities that are only valid on-shell in the open string/microscopic description. This will be made clear in Sections 4.4 and 4.5.

4. The Microscopic Formalism

4.1. Nekrasov’s technology

Nekrasov’s instanton technology^{16–20} is the crowning achievement of many years of developments in instanton calculus (see Ref. 23–25 and references therein). In a snapshot, this formalism allows to compute any integral over the moduli space of instantons in the chiral sector, for any value of the topological charge, and then to sum up exactly the instanton series (this series always have a finite radius of convergence). The remarkable mathematical property that underlies this result is that under certain conditions the integral over the instanton moduli space localizes over a finite number of field configurations that are labeled by colored partitions. A

generic formula thus looks like

$$\int dm^{(k)} \mathcal{O}(\mathbf{m}^{(k)}) e^{-\mathcal{S}_E} = q^k \sum_{|\vec{k}|=k} \mu_{\vec{k}}^2 \mathcal{O}_{\vec{k}}. \quad (29)$$

The integral in the left hand side is an integral over the moduli space of instantons of topological charge k , $\mathbf{m}^{(k)}$ denoting collectively the moduli. The expression $\mathcal{O}(\mathbf{m}^{(k)})$ corresponds to the operator \mathcal{O} , which can be an arbitrary chiral operator, evaluated on a particular instanton configuration labeled by $\mathbf{m}^{(k)}$. The right hand side of (29) is proportional to the k^{th} power of the instanton factor q , and is given by a finite sum over colored partitions \vec{k} of size $|\vec{k}| = k$. The factor $\mu_{\vec{k}}^2$ is a measure on the ensemble of colored partitions, and $\mathcal{O}_{\vec{k}}$ denotes the operator \mathcal{O} on the particular instanton configuration labeled by \vec{k} .

In order to get a better understanding of Eq. (29), let us give some more details on the ensemble of partitions and of colored partitions. An ordinary partition \mathbf{k} of size $|\mathbf{k}|$ is simply a decomposition of the positive integer $|\mathbf{k}|$ into a sum of positive integers. Thus ordinary partitions are in one-to-one correspondence with Young tableaux. For example, a partition of the integer 14 is depicted in Fig. 1. The number of boxes in the Young tableaux is equal to the topological charge. It is well-known that Young tableaux are in one-to-one correspondence with irreducible representations of the symmetric group with $|\mathbf{k}|$ elements, and thus it is rather natural to weight a given partition by the dimension $\dim R_{\mathbf{k}}$ of this irreducible representation (this is called the Plancherel measure in the mathematical literature). Indeed, a careful application of the localization techniques to the integral in Eq. (29) shows that the measure factor is given by

$$\mu_{\mathbf{k}} = \frac{1}{|\mathbf{k}|! \epsilon^{|\mathbf{k}|}} \dim R_{\mathbf{k}}, \quad (30)$$

where ϵ is a deformation parameter that we eventually take to zero (more is said on this deformation parameter in Section 4.3). The partitions that enter Eq. (29) are not ordinary partitions, but colored partitions, and thus the above discussion must be generalized. A colored partition is simply a collection of N ordinary partitions, $\vec{k} = (k_1, \dots, k_N)$. The integer N must coincide with the number of colors in the

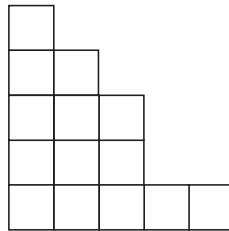


Fig. 1. The Young tableau associated with the partition $14 = 5 + 3 + 3 + 2 + 1$.

gauge theory. The measure factor $\mu_{\vec{k}}$ is then a natural generalization of the formula (30). A detailed discussion can be found for example in the Appendix of Ref. 7.

4.2. Instantons and $\mathcal{N} = 1$ gauge theories

A calculation based on instantons is non-perturbative, yet it is a priori valid only at weak coupling. In the case of $\mathcal{N} = 2$ supersymmetry, this is not a limitation. Indeed, $\mathcal{N} = 2$ gauge theories have a moduli space of vacua. The moduli space always contains a region where the theory is arbitrarily weakly coupled and the instanton approximation is exact. The full solution of the theory can then be obtained by *analytic continuation*. For these reasons, the instanton technology was originally developed with the aim of solving $\mathcal{N} = 2$ gauge theories.

The situation for theories with only $\mathcal{N} = 1$ supersymmetry is a priori much less favorable. Indeed, these theories generically do not have a moduli space, but a discrete set of vacua. *Some of these vacua are intrinsically strongly coupled* and thus a direct instanton analysis is impossible. In particular, observables can be given by fractional powers of the instanton factor, which is clearly incompatible with a direct instanton calculation.

The way this problem can be solved was explained in Ref. 6. The idea is to make the calculation in two steps. First, one considers *off-shell*, unphysical, correlators obtained by computing path integrals with fixed boundary conditions at infinity as in Eq. (7). By choosing appropriately \mathbf{a} (in such a way that $|a_i - a_j| \gg \Lambda$, where Λ is the dynamically generated scale of the theory), the path integrals can be forced to be weakly coupled, and the (unphysical) correlators are then given exactly in terms of instantons. Since the result must be holomorphic in \mathbf{a} , the value of the correlators for arbitrary \mathbf{a} can be obtained unambiguously by analytic continuation.

In the second step of the calculation, one computes the exact quantum superpotential for \mathbf{a} . We call this superpotential the microscopic superpotential $W_{\text{mic}}(\mathbf{a})$. Using the R-symmetry, W_{mic} can always be expressed in terms of the correlators of chiral operators, and can thus be computed exactly in the instanton approximation using step one. The superpotential W_{mic} has a fundamental property that distinguishes it from any other quantum superpotential previously used in the literature:⁶ *its critical points are in one-to-one correspondence with the full set of vacua of the theory*. In particular, the solutions corresponding to any number of cuts in the matrix model are obtained as critical points of a single superpotential, whereas in the matrix model approach one needs a different glueball superpotential $W_{\text{glue}}^{(r)}$ for each value of r (r corresponds to the number of cuts), as explained in Section 2. The fact that all the vacua are found as extrema of a single superpotential W_{mic} is of course the signature of the microscopic nature of our analysis.

We can now understand how the strongly coupled $\mathcal{N} = 1$ vacua are described with the help of instantons, in an indirect way. We first compute \mathbf{a} -dependent unphysical correlators and $W_{\text{mic}}(\mathbf{a})$ in the instanton approximation. The corresponding instanton series have a finite radius of convergence. We then solve Eq. (10). Some

of the solutions $\mathbf{a} = \mathbf{a}^*$ lie inside the disk of convergence of the instanton series. These solutions correspond to weakly coupled, Coulomb-like vacua, for which all the physical correlators can be expanded as series in q and could actually be computed directly with instantons without using our two-step procedure. In addition, it turns out that there are also solutions $\mathbf{a} = \mathbf{a}^*$ to Eq. (10) that lie outside the disk of convergence of the instanton series (they are typically on the boundary of this disk). *These solutions correspond to the strongly coupled vacua.* Expanding around such strongly coupled solutions automatically produce series containing fractional powers of the instanton factor. For example, this is how the gluino condensate $\sim q^{1/N}$ is obtained in the pure $\mathcal{N} = 1$ gauge theory.

4.3. Important technicalities

- Physically, the deformation parameter ϵ (the so-called Ω -background) appearing in Eq. (30) provides a nice IR regulator. The calculation of the scalar correlators $[\text{Tr } X^k]$ in the $\epsilon \rightarrow 0$ limit goes essentially as in the $\mathcal{N} = 2$ theory.²¹ The glueball correlators $[\text{Tr } W^\alpha W_\alpha X^k]$, which are zero in the $\mathcal{N} = 2$ case, are much more interesting. Eq. (8) shows that they are related to the next-to-leading order in the small ϵ expansion of the scalars.^{7,8}
- Some correlators turn out to be ambiguous in instanton calculus. For example, if $\text{Tr } X^k$ for $k \geq 2N$ is inserted in the path integral, the result is typically 0/0. This singular behavior is due to the small instanton singularities on the instanton moduli space. The discussion in Section 3.2 gives a clear physical interpretation of these ambiguities: they correspond to the ambiguities in the definitions (27) of the variables. Each set of definitions thus corresponds to a choice of regularization of the instanton moduli space. In Nekrasov's formalism, we use the non-commutative deformation of the theory to regularize the moduli space. This particular scheme yields the canonical definition of Theorem 3.1.

An interesting extension of these ideas is as follows. The non-commutative deformation does not work for all gauge groups, but we conjecture that *there always exists a unique canonical regularization of the instanton moduli space, for all gauge groups, corresponding to a choice of variables that make the quantum corrections in the anomaly equations implicit.*

4.4. The duality between the colored partitions and the matrix model formalisms

We have developed a new “open string” formalism to solve $\mathcal{N} = 1$ theories, based on first-principle path integral calculations. We also have at our disposal the closed string point of view, based on summing over hermitian matrices and the Dijkgraaf-Vafa glueball superpotential. The two formalisms are clearly completely different, in spite of some formal similarities when one exchange scalar and glueball operators, colored partitions and matrices, identities and equations of motion.⁸

The fundamental result, which is mathematically summarized by Eq. (14), is that when both formalisms are *on-shell*, they yield the same correlators. Let us give a few more details on how this equivalence works.

In the microscopic formalism, we first compute the generating functions $R_{\text{mic}}(z; \mathbf{a})$ and $S_{\text{mic}}(z; \mathbf{a})$, whereas in the matrix model formalism one deals with different generating functions $R_{\text{MM}}(z; \mathbf{s})$ and $S_{\text{MM}}(z; \mathbf{s})$. In the matrix model formalism, extremizing the glueball superpotential is equivalent to the quantization of the periods of $R_{\text{MM}}dz$. This means that

$$dW_{\text{glue}} \sim \oint R_{\text{MM}} dz \mod 2i\pi\mathbb{Z}. \quad (31)$$

We have explained in Section 3 that the condition $dW_{\text{glue}} = 0$ ensures the consistency with the open string formulation of the theory (this is essentially Theorem 3.2). This means that in the microscopic approach, which is based on the open string formulation and thus in which the constraints (27) are trivially satisfied, the conditions derived from $dW_{\text{glue}} = 0$ should correspond to identities valid off-shell. This is exactly what is found: one can easily show that

$$\oint R_{\text{mic}}(z; \mathbf{a}) dz \in 2i\pi\mathbb{Z}, \quad (32)$$

for any \mathbf{a} . Similarly, in the matrix model formalism,

$$\oint S'_{\text{MM}}(z; \mathbf{s}) dz = 0 \quad (33)$$

is a trivial identity valid off-shell (for any \mathbf{s}). It follows from the loop equations of the matrix model. On the other hand, S'_{mic} does not satisfy an identity like (33) for any \mathbf{a} , but rather we have

$$dW_{\text{mic}} \sim \oint S'_{\text{mic}} dz. \quad (34)$$

Consistency between the formalisms thus comes from the fact that identities in one formalism, like (32) and (33), are exchanged with equations of motion in the other formalism, (31) and (34). Eventually, the full equivalence when $\mathbf{a} = \mathbf{a}^*$ and $\mathbf{s} = \mathbf{s}^*$, i.e. when $dW_{\text{mic}} = 0$ and $dW_{\text{glue}} = 0$, can be proven.⁸

4.5. The anomaly equations

We can now provide a full non-perturbative discussion of the anomaly equations and of Theorem 3.1. As in any first-principle, microscopic approach, the anomaly polynomials are expressed as variations of the microscopic quantum effective action. In the chiral sector we are discussing, this reduces to variations of the microscopic superpotential. If we denote by

$$\mathcal{A}_n(\mathbf{a}) = -N \sum_{k \geq 0} gk u_{n+k+1, \text{mic}} + 2 \sum_{k_1+k_2=n} u_{k_1, \text{mic}} v_{k_2, \text{mic}} \quad (35)$$

$$\mathcal{B}_n(\mathbf{a}) = -N \sum_{k \geq 0} gk v_{n+k+1, \text{mic}} + \sum_{k_1+k_2=n} v_{k_1, \text{mic}} v_{k_2, \text{mic}} \quad (36)$$

the anomaly polynomials in the microscopic formalism (compare with Eq. (22) and (24)), one can construct first order partial differential operators \mathcal{L}_n and \mathcal{J}_n ,^{7,8} that are the quantum versions of the operators L_n and J_n discussed in Section 3, such that

$$\mathcal{L}_n \cdot W_{\text{mic}} = \mathcal{A}_n, \quad \mathcal{J}_n \cdot W_{\text{mic}} = \mathcal{B}_n. \quad (37)$$

More precisely, the operators \mathcal{L}_n and \mathcal{J}_n have the form

$$\mathcal{L}_n = \sum_{i=1}^N \ell_{n,i}(\mathbf{a}) \frac{\partial}{\partial a_i}, \quad \mathcal{J}_n = \sum_{i=1}^N j_{n,i}(\mathbf{a}) \frac{\partial}{\partial a_i}, \quad (38)$$

where $\ell_{n,i}(\mathbf{a})$ and $j_{n,i}(\mathbf{a})$ are given in terms of period integrals of the one-forms $z^{n+1}R_{\text{mic}}dz$ and $z^{n+1}S_{\text{mic}}dz$ respectively.^{7,8}

One can then straightforwardly compute, for example, the quantum corrections to the perturbative super-Virasoro algebra given by (18) and (23). One then finds in particular that the quantum algebra does not close on the operators \mathcal{L}_n and \mathcal{J}_n alone.

We can also compute explicitly the action of \mathcal{L}_n and \mathcal{J}_n on the variables u_m and v_m . We now understand why the simple ansatz (28) cannot work: $\mathcal{L}_n \cdot u_{m,\text{mic}}$ is a perfectly well defined function of \mathbf{a} , but \mathbf{a} itself is a multi-valued function of the $u_{p,\text{mic}}$ (because of monodromies in the \mathbf{a} -plane, that are strictly similar to the familiar monodromies in the Seiberg-Witten moduli space of $\mathcal{N} = 2$ theories). Thus $\mathcal{L}_n \cdot u_{m,\text{mic}}$ cannot possibly be expressed as a polynomial in the u_p s.

5. Conclusions

We have developed a first-principle, text-book like approach to all the known exact results in $\mathcal{N} = 1$ super Yang-Mills theories. This provides in particular an explicit and exactly solvable example of the open/closed string duality. The duality translates mathematically into a beautiful equivalence between formalisms based on sums over colored partitions on the one hand and on matrix integrals on the other hand. It is a unique case where both sides of the duality are now exactly solved.

Many generalizations are possible. Particularly interesting examples include turning on some backgrounds (corresponding to higher genus in the matrix model for instance). Each case must yield a non-trivial and beautiful equivalence between the open and closed string formulations.

Acknowledgments

This work is supported in part by the belgian Fonds de la Recherche Fondamentale Collective (grant 2.4655.07), the belgian Institut Interuniversitaire des Sciences Nucléaires (grant 4.4505.86), the Interuniversity Attraction Poles Programme (Belgian Science Policy) and by the European Commission FP6 programme MRTN-CT-2004-005104 (in association with V. U. Brussels).

References

1. R. Gopakumar and C. Vafa, *Adv. Theor. Math. Phys.* **3** (1999) 1415, hep-th/9811131, C. Vafa, *J. Math. Phys.* **42** (2001) 2798, hep-th/0008142, F. Cachazo, K. Intriligator and C. Vafa, *Nucl. Phys. B* **603** (2001) 3, hep-th/0103067.
2. R. Dijkgraaf and C. Vafa, *Nucl. Phys. B* **644** (2002) 3, hep-th/0206255; *Nucl. Phys. B* **644** (2002) 21, hep-th/0207106; *A Perturbative Window into Non-Perturbative Physics*, hep-th/0208048.
3. F. Ferrari, *Nucl. Phys. B* **648** (2002) 161, hep-th/0210135.
4. R. Dijkgraaf, M.T. Grisaru, C.S. Lam, C. Vafa and D. Zanon, *Phys. Lett. B* **573** (2003) 138, hep-th/0211017.
5. F. Cachazo, M.R. Douglas, N. Seiberg and E. Witten, *JHEP* **12** (2002) 071, hep-th/0211170.
6. F. Ferrari, *JHEP* **10** (2007) 065, arXiv:0707.3885.
7. F. Ferrari, S. Kuperstein and V. Wens, *JHEP* **10** (2007) 101, arXiv:0708.1410.
8. F. Ferrari, *JHEP* **11** (2007) 001, arXiv:0709.0472.
9. F. Ferrari, *Nucl. Phys. B* **770** (2007) 371, hep-th/0701220.
10. F. Ferrari and V. Wens, Consistency conditions in the chiral ring of super Yang-Mills theories, to appear in *Nucl. Phys. B* (in press).
11. F. Ferrari, *Phys. Rev. D* **67** (2003) 85013, hep-th/0211069.
12. F. Ferrari, *Phys. Lett. B* **557** (2003) 290, hep-th/0301157.
13. F. Cachazo, N. Seiberg and E. Witten, *JHEP* **02** (2003) 042, hep-th/0301006.
14. F. Cachazo, N. Seiberg and E. Witten, *JHEP* **04** (2003) 018, hep-th/0303207.
15. N. Seiberg and E. Witten, *Nucl. Phys. B* **426** (1994) 19, erratum **B 430** (1994) 485, hep-th/9407087; *Nucl. Phys. B* **431** (1994) 484, hep-th/9408099.
16. N. Nekrasov, *Adv. Theor. Math. Phys.* **7** (2004) 831, hep-th/0206161.
17. N. Nekrasov, Seiberg-Witten Prepotential from Instanton Counting, in *Proc. International Congress of Mathematicians (ICM 2002)*, hep-th/0306211.
18. N. Nekrasov and A. Okounkov, Seiberg-Witten Theory and Random Partitions, hep-th/0306238.
19. N. Nekrasov and S. Shadchin, *Comm. Math. Phys.* **252** (2004) 359, hep-th/0404225.
20. A. Marshakov and N. Nekrasov, *JHEP* **01** (2007) 104, hep-th/0612019.
21. F. Fucito, J.F. Morales, R. Poghossian and A. Tanzini, *JHEP* **01** (2006) 031, hep-th/0510173.
22. H. Itoyama and K. Maruyoshi, *Phys. Lett. B* **650** (2007) 298, arXiv:0704.1060; *Nucl. Phys. B* **796** (2008) 246, arXiv:0710.4377.
23. N. Dorey, V.V. Khoze and M.P. Mattis, *Phys. Rev. D* **54** (1996) 2921, hep-th/9603136, F. Fucito and G. Travaglini, *Phys. Rev. D* **55** (1997) 1099, hep-th/9605215, N. Dorey, V.V. Khoze and M.P. Mattis, *Phys. Rev. D* **54** (1996) 7832, hep-th/9607202, N. Dorey, V.V. Khoze and M.P. Mattis, *Phys. Lett. B* **396** (1997) 141, hep-th/9612231, V.V. Khoze, M.P. Mattis and M.J. Slater, *Nucl. Phys. B* **536** (1998) 69, hep-th/9804009.
24. D. Bellisai, F. Fucito, A. Tanzini and G. Travaglini, *Phys. Lett. B* **480** (2000) 365, hep-th/0002110, D. Bellisai, F. Fucito, A. Tanzini and G. Travaglini, *JHEP* **07** (2000) 017, hep-th/0003272, N. Dorey, T.J. Hollowood and V.V. Khoze, *JHEP* **03** (2001) 040, hep-th/0011247, F. Fucito, J.F. Morales and A. Tanzini, *JHEP* **07** (2001) 012, hep-th/0106061.
25. T.J. Hollowood, *JHEP* **03** (2002) 038, hep-th/0201075, T.J. Hollowood, *Nucl. Phys. B* **639** (2002) 66, hep-th/0202197.